Chapter 4

An Application to Simulation

4.1. Introduction

In Sections 5.9 and 10.4.4 of the book we showed how heavy-traffic stochastic-process limits for queues can be used to help plan queueing simulations. In this chapter we discuss another application of stochastic-process limits to simulation. We draw on Glynn and Whitt (1992a). In Section 4.2 we show how stochastic-process limits and the continuous-mapping approach can be used to determine general criteria for sequential stopping rules to be asymptotically valid.

Yet another application of stochastic-process limits and the continuous-mapping approach to simulation is contained in Glynn and Whitt (1992b). Glynn and Whitt (1992b) shows how stochastic-process limits and the continuous-mapping approach can be exploited to determine the asymptotic efficiency of simulation estimators. These two applications can be applied to queueing simulations, but they are not limited to queueing simulations.

4.2. Sequential Stopping Rules for Simulations

In this section, following Glynn and Whitt (1992a), we show how FCLTs and the continuous-mapping approach can be used to establish general conditions for the asymptotic validity of sequential stopping rules for stochastic simulations. The general conditions are expressed in terms of FCLTs and FWLLNs. The conditions allow the possibility of limit processes with discontinuous sample paths, but usually the limit process will be related to

Brownian motion, and thus have continuous sample paths. We use the composition and inverse maps to demonstrate the asymptotic validity.

The goal is to estimate a deterministic parameter $\alpha \in \mathbb{R}^k$. We start with an $\mathbb{R}^k$-valued stochastic process $Y \equiv \{Y(t) : t > 0\}$ called the *estimation process*. We think of $Y(t)$ as being the estimate of α based on a simulation with runlength t. The results also apply to statistical estimation more generally, but we are especially concerned with simulation.

With simulation, a common problem is to estimate a steady-state mean vector α. The simulation may be used to generate a stochastic process $X \equiv \{X(t) : t \geq 0\}$, where $X(t) \Rightarrow X(\infty)$ in $\mathbb{R}^k$ as $t \to \infty$. We may then want to estimate the steady-state mean $\alpha \equiv EX(\infty) \equiv [EX^1(\infty), \ldots, EX^k(\infty)]$ by the sample mean

$$Y(t) \equiv t^{-1} \int_0^t X(s)ds, \quad t > 0 \ . \tag{2.1}$$

That is a common way for the estimation process Y to arise, but not the only way.

The simulator must select a runlength t. The runlength can be selected either in advance or sequentially while the simulation is in process. The principal disadvantage of selecting the runlength in advance is that the posterior precision of the estimator may not be appropriate. Since the volume of the confidence set (the width of a confidence interval in one dimension) is unknown in advance, the volume may be too large to be of practical use (meaning that the preassigned runlength was too small) or too small (meaning that computational resources were wasted in refining the estimator beyond the level of accuracy required).

We are interested in sequential procedures in which we let the simulation run until the volume of a confidence set achieves a prescribed value. That avoids the problems associated with preassigned runlengths, but new difficulties are introduced because the runlength is now randomly determined. The first difficulty is that we no longer have direct control of the amount of simulation time to be generated or the amount of computer time to be expended. Consequently, the runlength may turn out to be much longer than we want. On the other hand, it is possible that the runlength may turn out to be inappropriately short. This creates certain statistical difficulties that can compromise the accuracy of such procedures. For example, it is known that in many statistical settings, the point estimator and the variance estimator are positvely correlated. Since the volume of a confidence set is typially determined by the variance estimator, this suggests that the confidence set volume will tend to be small when the point estimator is small.

Consequently, the resulting sequential procedure will tend to terminate early in situations in which the point estimator is too small, leading to possibly significant coverage problems for the confidence sets. Nevertheless, sequential stopping rules are of interest because of the possiblity of automatically obtaining prescribed precision.

Various sequential stopping rules for simulation estimators have been proposed and investigated empirically. Among these are sequential procedures involving: batch means in Law and Carson (1979) and Law and Kelton (1982), regenerative simulation in Fishman (1977) and Lavenberg and Sauer (1977) and spectral methods in Heidelberger and Welch (1981a, b, 1983); see pages 81, 92, 97 and 103 of Bratley, Fox and Schrage (1987) for an overview. Unfortunately, however, the empirical evidence is not entirely encouraging. Evidently, care must be taken in the design and implementation of sequential procedures to avoid inappropriate early termination. On the positive side, the sequential procedures do tend to perform well when the run lengths are relatively long, which is achieved in part by having a suitably small precribed volume for the confidence set. The observed good performance with small prescribed confidence set volumes is consistent with the asymptotic theory to be developed below. The asymptotic theory for general simulation estimators below is in turn consistent with the classical asymptotic theory associated with the sample mean of i.i.d. random variables; we cite references below.

4.2.1. The Mathematical Framework

To start, we assume that the estimation process Y satisfies a CLT, i.e.,

$$\phi(t)[Y(t)-\alpha] \Rightarrow \Gamma L \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad t \to \infty\,, \tag{2.2}$$

where Γ is a nonsingular $k \times k$ scaling matrix and $\phi(t)$ is a real-valued scaling function with $\phi(t) \to \infty$ as $t \to \infty$. The common case for ϕ is $\phi(t) = t^{1/2}$, in which case the limit L in (2.2) typically is $N(0, I)$, a standard normal random vector with the identity matrix I as its covariance matrix, but we want to allow for other possibilities. With heavy-tailed probability distributions or long-range dependence, we might have $\phi(t) = t^\gamma$ for $\gamma < 1/2$ or, more generally, ϕ regularly varying with index γ. The treatment here generalizes Glynn and Whitt (1992a) by allowing regularly varying scaling functions instead of simple powers.

As a consequence of (2.2),

$$Y(t) \Rightarrow \alpha \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad t \to \infty\,. \tag{2.3}$$

The limit (2.3) says that the estimation process is *weakly consistent.* Of course, weak consistency is a minimal requirement.

We assume that the confidence sets are all based on a bounded measurable subset A of $\mathbb{R}^k$ with $m(A) > 0$, where m is Lebesgue measure on $\mathbb{R}^k$. To obtain approximate $100(1-\delta)\%$ confidence sets for α, we assume that

$$P(L \in A) = 1 - \delta \quad \text{and} \quad P(L \in \partial A) = 0 \ , \tag{2.4}$$

where L is the limiting random variable in (2.2) and ∂A is the boundary of the set A, i.e., $\partial A = A^- - A^o$, where A^- and A^o are the closure and interior of A. Given that we know A and Γ, we can let the *confidence set* be

$$\tilde{C}(t) \equiv Y(t) - \phi(t)\Gamma A \ , \tag{2.5}$$

where

$$z + QA \equiv \{x \in R^d : \text{there exists } y \in A \text{ such that } x = z + Qy\} \, .$$

The confidence set $\tilde{C}(t)$ in (2.5) clearly depends on t. When the runlength t is specified in advance, the confidence set is asymptotically valid, in the sense of the following proposition.

Proposition 4.2.1. *If* (2.2) *and* (2.4) *hold, then*

$$P(\alpha \in \tilde{C}(t)) \to 1 - \delta \quad \textit{as} \quad t \to \infty$$

for $\tilde{C}(t)$ *in* (2.5).

Proof. Since Γ is nonsingular,

$$P(\alpha \in \tilde{C}(t)) = P(\Gamma^{-1}\phi(t)(Y(t) - \alpha) \in A) \ ,$$

but

$$\Gamma^{-1}\phi(t)(Y(t) - \alpha) \Rightarrow \Gamma^{-1}\Gamma L = L \quad \text{as} \quad t \to \infty$$

by (2.2). Since (2.4) holds,

$$P(\Gamma^{-1}\phi(t)(Y(t) - \alpha) \in A) \to P(L \in A) = 1 - \delta \quad \text{as} \quad t \to \infty$$

by Theorem 11.3.4 (v) in the book. ■

Of course, in applications the scaling matrix Γ is typically unknown, so that it too must be estimated. We assume that there is an estimator $\Gamma(t)$ that is weakly consistent, i.e.,

$$\Gamma(t) \Rightarrow \Gamma \quad \text{in} \quad \mathbb{R}^{k^2} \quad \text{as} \quad t \to \infty \, . \tag{2.6}$$

Given an estimator $\Gamma(t)$, $t > 0$, we can form approximate confidence sets based on $\Gamma(t)$. For that purpose, let

$$C(t) \equiv Y(t) - \phi(t)\Gamma(t)A \ . \tag{2.7}$$

We now extend Proposition 4.2.1 to include $\Gamma(t)$ instead of Γ.

Proposition 4.2.2. *If, in addition to the assumptions of Proposition* 4.2.1 *above,* (2.6) *holds, then*

$$P(\alpha \in C(t)) \to 1 - \delta \quad as \quad t \to \infty$$

for $C(t)$ *in* (2.7).

Proof. By (2.2) above and Theorem 11.4.5 in the book,

$$(\Gamma(t), \phi(t)(Y(t) - \alpha)) \Rightarrow (\Gamma, \Gamma L) \quad \text{as} \quad t \to \infty \ .$$

Then noting that matrix inversion is continuous at all nonsingular limits, we can deduce that $\Gamma(t)$ is nonsingular, and thus invertible, for all sufficiently large t and then apply the continuous mapping theorem to obtain

$$\Gamma(t)^{-1}\phi(t)(Y(t) - \alpha) \Rightarrow \Gamma^{-1}\Gamma L \quad \text{as} \quad t \to \infty \, .$$

The rest of the proof is the same as the last part of the proof of Proposition 4.2.1. ■

We now use the confidence set $C(t)$ in (2.7) to define sequential stopping rules. Recall that, for a generic (measurable) set $B \subseteq \mathbb{R}^k$, $m(B)$ denotes the k-dimensional volume (Lebesgue measure) of the set. Of course, when $k = 1$ and B is an interval, $m(B)$ is just the length of the interval. We first consider the case in which the procedure terminates when the k^{th} root of the volume of the confidence region $C(t)$ drops below a prescribed level ϵ. [It is natural to use the k^{th} root, because $m(cB)^{1/k} = cm(B)^{1/k}$ for a scalar c.] We call such a procedure an *absolute-precision sequential stopping rule.* For such a rule, the time $\tilde{T}(\epsilon)$ at which the simulation terminates execution is defined by

$$\tilde{T}(\epsilon) = \inf\{t \ge 0 : m(C(t))^{1/k} < \epsilon\} \ . \tag{2.8}$$

Actually, this stopping rule needs to be modified, because $\tilde{T}(\epsilon)$ in (2.8) can terminate much too early if the estimator $\Gamma(t)$ is badly behaved for small t. To see this, suppose that $P(\Gamma(1) = 0, \, m(C(t)) = 1, \, 0 \le t < 1) = 1$. In this case, $\tilde{T}(\epsilon) = 1$ for $\epsilon < 1$, so $C(\tilde{T}(\epsilon)) = Y(1)$ for $\epsilon < 1$. Hence, in this example, $P(\alpha \in C(\tilde{T}(\epsilon))) = P(\alpha = Y(1))$ for $\epsilon < 1$. Hence convergence of the coverage probability of the region $C(T(\epsilon))$ to the nominal level $1-\delta$ does *not* occur when we let $\epsilon \downarrow 0$.

In order for the asymptotic theory to be relevant to the sequential stopping problem, it is necessary that $T(\epsilon) \to \infty$ as $\epsilon \downarrow 0$. In other words, small values of the precision constant ϵ need to correspond to large values of simulation time. We can force the termination time to behave in this way if we inflate the volume $m(C(t))$ slightly. Let $a(t)$ be a strictly positive function that decreases monotonically to 0 as $t \to \infty$ and satisfies $a(t) = o(\phi(t))$ as $t \to \infty$, where ϕ is the scaling function in the CLT (2.2). Then set

$$T_1(\epsilon) \equiv inf\{t \geq 0 : m(C(t))^{1/k} + a(t) < \epsilon\} . \tag{2.9}$$

Note that

$$T_1(\epsilon) \geq t_1(\epsilon) \equiv \inf\{t \geq 0 : a(t) < \epsilon\} \to \infty \quad \text{as} \quad \epsilon \downarrow 0 . \tag{2.10}$$

Thus the early termination associated with $\tilde{T}(\epsilon)$ in (2.8) is prevented by incorporating the deterministic function $a(t)$ in $T_1(\epsilon)$ in (2.9). For practical purposes, it remains to determine appropriate functions $a(t)$, though.

An alternative to the absolute-precision sequential stopping rule in (2.9) is a *relative-precision sequential stopping rule*. The basic idea here is that the simulation should terminate when the k^{th} root of the volume of the confidence region is less than an ϵ^{th} fraction of the norm of the parameter α, denoted by $\|\alpha\|$, under the additional condition that $\|\alpha\| > 0$. Since $Y(t)$ is an estimator for α, this suggests replacing $T_1(\epsilon)$ with

$$T_2(\epsilon) = \inf\{t \geq 0 : m(C(t))^{1/k} + \alpha(t) < \epsilon \|Y(t)\|\} . \tag{2.11}$$

The question now is: When are these sequentially stopping rules asymptotically valid? That is, when can we conclude that

$$P(\alpha \in C(T(\epsilon))) \to 1 - \delta \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.12}$$

for $T(\epsilon)$ being $T_1(\epsilon)$ in (2.9) or $T_2(\epsilon)$ in (2.11)?

It turns out that, unlike in Propositions 4.2.1 and 4.2.2, the assumed convergence in (2.2) and (2.6) is *not* enough to achieve asymptotic validility for the sequential stopping rules. That is for the same reason that CLTs involving random time change require extra conditions. However, we do obtain asymptotic validity if we replace the ordinary CLT in (2.2) by a FCLT and if we replace the ordinary WLLN in (2.6) by a SLLN or FWLLN.

(Recall that the SLLN implies a FSLLN by Corollary 3.2.1 in Chapter 3 here, which in turn implies a FWLLN, so that the SLLN is the stronger condition.)

For that purpose, we form scaled processes indexed by ϵ in the function space $D((0,\infty),\mathbb{R}^k)$. We work with time domain $(0,\infty)$ instead of $[0,\infty)$ in order to avoid having to deal with possible singularities in the estimation process Y at the origin $t=0$. For example, such singularities occur in the special case in (2.1). Recall that $x_n \to x$ in $D((0,\infty),\mathbb{R}^d)$ if the restrictions converge in $D([t_0,t_1],\mathbb{R}^d)$ for all t_0, t_1 with $0 < t_0 < t_1 < \infty$.

Given the estimation process Y, the associated scaled estimation processes are

$$\mathbf{Y}_\epsilon(t) \equiv \phi(\epsilon^{-1})[Y(t/\epsilon) - \alpha], \quad t > 0\,. \tag{2.13}$$

For the results below we need to assume that the scaling function ϕ in (2.13) is regularly varying with index γ, denoted by $\phi \in \mathcal{R}(\gamma)$; see Appendix A in the book. We also assume that ϕ is a homeomorphism of $\mathbb{R}^+$, which implies that $\phi(0) = 0$ and ϕ is strictly increasing.

4.2.2. The Absolute-Precision Sequential Estimator

We first state a result for the absolute-precision sequential estimator $T_1(\epsilon)$ in (2.9).

Theorem 4.2.1. *Let $D \equiv D((0,\infty),\mathbb{R}^k)$ be endowed with the WM_2 or any other Skorohod topology. Suppose that*

$$\mathbf{Y}_\epsilon \Rightarrow \Gamma\mathbf{Z} \quad in \quad D \quad as \quad \epsilon \downarrow 0\,, \tag{2.14}$$

for $\mathbf{Y}_\epsilon$ in (2.13), where (2.4) holds with $L = \mathbf{Z}(1)$, $P(t \in Disc(\mathbf{Z})) = 0$ for all t, ϕ is a homeomorphism of $\mathbb{R}_+$, $\phi \in \mathcal{R}(\gamma)$ for $\gamma > 0$, and Γ is nonsingular. If, in addition,

$$\Gamma(t) \to \Gamma \quad w.p.1 \quad in \quad \mathbb{R}^{k^2} \quad as \quad t \to \infty\,, \tag{2.15}$$

then as $t \to \infty$ or as $\epsilon \downarrow 0$

(a) $\phi(t)[m(C(t))^{1/k} + a(t)] \to m(\Gamma A)^{1/k}$ *w.p.1,*

(b) $\epsilon\phi(T_1(\epsilon)) \to m(\Gamma A)^{1/k}$ *w.p.1,*

(c) $\epsilon^{-1} m(C(T_1(\epsilon)))^{1/k} \to 1$ *w.p.1,*

(d) $\epsilon^{-1}[Y(T_1(\epsilon)) - \alpha] \Rightarrow m(\Gamma A)^{-1/k}\Gamma\mathbf{Z}(1) \quad in \quad \mathbb{R}^k$,

(e) $P(\alpha \in C(T_1(\epsilon))) \to 1 - \delta$ *(asymptotic validity).*

In our proof of Theorem 4.2.1, we use the following lemma, which shows the scaling implications for the limit process $\mathbf{Z}$ from having the FCLT in (2.14) hold with the regularly varying scaling function ϕ in (2.13). The result is a consequence of Theorem 5.2.1 in the book, but we give a direct proof here.

Lemma 4.2.1. *If the FCLT* (2.14) *holds with* $\phi \in \mathcal{R}(\gamma)$, $\gamma > 0$, *for* ϕ *in* (2.13), *then*

$$\{\mathbf{Z}(ct) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{c^{-\gamma}\mathbf{Z}(t) : t \geq 0\} \qquad (2.16)$$

for any $c > 0$.

Proof. Note that $\mathbf{Y}_\epsilon \circ ce \Rightarrow \mathbf{Z} \circ ce$ as $\epsilon \downarrow 0$. On the other hand,

$$\mathbf{Y}_\epsilon \circ c\mathbf{e} = \frac{\phi(\epsilon^{-1})}{\phi(ce^{-1})}\mathbf{Y}_{\epsilon/c} \Rightarrow c^{-\gamma}\mathbf{Z} \quad \text{as} \quad \epsilon \downarrow 0\,, \qquad (2.17)$$

using the regular variation to get $\phi(\epsilon^{-1})/\phi(c/\epsilon^{-1}) \to c^{-\gamma}$ as $\epsilon \downarrow 0$ for every $c > 0$; see Appendix A in the book. ∎

Proof of Theorem 4.2.1. (a) Let

$$V(t) \equiv m(C(t))^{1/k} + a(t), \quad t > 0\ . \qquad (2.18)$$

By the spatial invariance and scaling properties of Lebesgue measure m on $\mathbb{R}^k$,

$$\begin{aligned} m(C(t))^{1/k} &= m(Y(t) - \phi(t)^{-1}\Gamma(t)A)^{1/k} \\ &= m(-\phi(t)^{-1}\Gamma(t)A)^{1/k} = \phi(t)^{-1}m(\Gamma(t)A)^{1/k}\ . \end{aligned} \qquad (2.19)$$

Since A is a bounded set, $\Gamma(t)A$ is contained in a bounded set for all sufficiently large t w.p.1. Thus, we can apply the bounded convergence theorem to deduce that

$$m(\Gamma(t)A)^{1/k} \to m(\Gamma A)^{1/k} \quad \text{w.p.1 as} \quad t \to \infty\,. \qquad (2.20)$$

Since Γ is nonsingular, $m(A) > 0$ and $a(t) = o(\phi(t)^{-1})$ as $t \to \infty$, (2.18) and (2.20) imply that

$$\phi(t)V(t) \to m(\Gamma A)^{1/k} > 0 \quad \text{w.p.1 as} \quad t \to \infty\,. \qquad (2.21)$$

(b) By the definition of $T_1(\epsilon)$ in (2.9), $V(T_1(\epsilon)-1) \geq \epsilon$ and there exists a random variable $Z(\epsilon)$ with $0 \leq Z(\epsilon) \leq 1$ such that $V(T_1(\epsilon)+Z(\epsilon)) < \epsilon$. (Note that $V(t)$ is not necessarily monotone.) By (2.21) and the fact that $T_1(\epsilon) \to \infty$ w.p.1 as $\epsilon \downarrow 0$,

$$\limsup_{\epsilon\downarrow 0} \epsilon\phi(T_1(\epsilon)) \leq \limsup_{\epsilon\downarrow 0} \phi(T_1(\epsilon))V(T_1(\epsilon)-1) = m(\Gamma A)^{1/k} \quad \text{w.p.1} \tag{2.22}$$

and

$$\liminf_{\epsilon\downarrow 0} \epsilon\phi(T_1(\epsilon)) \geq \liminf_{\epsilon\downarrow 0} \phi(T_1(\epsilon))(V(T_1(\epsilon))+Z(\epsilon)) = m(\Gamma A)^{1/k} \quad \text{w.p.1.} \tag{2.23}$$

(c) Note that

$$m(C(T_1(\epsilon))^{1/k} = \phi(T_1(\epsilon))^{-1} m(\Gamma(T_1(\epsilon))A)^{1/k} \tag{2.24}$$

and recall that $m(\Gamma(t)A) \to m(\Gamma A)$ w.p.1 as $t \to \infty$, so that $m(\Gamma(T_1(\epsilon)) \to m(\Gamma A)$ w.p.1 as $\epsilon \downarrow 0$. By (b), $\epsilon^{-1}\phi(T_1(\epsilon)) \to m(\Gamma A)^{-1/k}$. Hence

$$\begin{aligned} \epsilon^{-1} m(C(T_1(\epsilon)))^{1/k} &= \epsilon^{-1}\phi(T_1(\epsilon))^{-1} m(\Gamma(T_1(\epsilon)A)^{1/k} \\ &\to m(\Gamma A)^{-1/k} m(\Gamma A)^{1/k} = 1 \text{ w.p.1} \quad \text{as} \quad \epsilon \downarrow 0. \end{aligned} \tag{2.25}$$

(d) From the assumed FCLT (2.14), $\mathbf{Z}_\epsilon \Rightarrow \Gamma\mathbf{Z}$ in $D((0,\infty),M_2)$ as $\epsilon \downarrow 0$, where

$$\mathbf{Z}_\epsilon(t) \equiv \mathbf{Y}_{1/\phi^{-1}(\epsilon^{-1})}(t) \equiv \epsilon^{-1}(Y(\phi^{-1}(\epsilon^{-1})t) - \alpha), \quad t > 0\,. \tag{2.26}$$

Now form the deterministic function

$$\boldsymbol{\psi}_\epsilon(t) = \frac{\phi^{-1}(\epsilon^{-1}t)}{\phi^{-1}(\epsilon^{-1})}, \quad t > 0\,. \tag{2.27}$$

Since ϕ is a homeomorphism of $\mathbb{R}_+$, the inverse ϕ^{-1} exists and is itself an homeomorphism of $\mathbb{R}_+$. Moreover, since $\phi \in \mathcal{R}(\gamma)$, $\phi^{-1} \in \mathcal{R}(\gamma^{-1})$ by Theorem 1.5.12 of Bingham, Goldie and Tengels (1989). Hence

$$\boldsymbol{\psi}_\epsilon \to \mathbf{e}^{1/\gamma} \quad \text{in} \quad D \quad \text{as} \quad \epsilon \downarrow 0\,. \tag{2.28}$$

We can apply the continuous-mapping theorem with the composition map taking $D \times D$ into D with (2.26)–(2.28), using Theorem 13.2.3 in the book, to conclude that

$$\mathbf{Z}'_\epsilon \Rightarrow \Gamma\mathbf{Z} \circ \mathbf{e}^{1/\gamma} \text{ in } (D, M_2) \text{ as } \epsilon \downarrow 0\,, \tag{2.29}$$

where

$$\mathbf{Z}'_\epsilon(t) \equiv (\mathbf{Z}_\epsilon \circ \boldsymbol{\psi}_\epsilon)(t) \equiv \epsilon^{-1} Y(\phi^{-1}(\epsilon^{-1}t) - \alpha), \quad t > 0\,. \tag{2.30}$$

Finally, we can apply the continuous-mapping theorem with the composition map taking $D((0,\infty),\mathbb{R}^k) \times \mathbb{R}$ into $\mathbb{R}^k$ with (2.30), invoking Proposition 13.2.1 in the book and part (b) here, to obtain

$$\epsilon^{-1} Y(T_1(\epsilon) - \alpha) = \mathbf{Z}'_\epsilon(\epsilon\phi(T_1(\epsilon)) \Rightarrow \Gamma(\mathbf{Z} \circ \mathbf{e}^{1/\gamma})(m(\Gamma A)^{1/k}) \quad \text{in} \quad \mathbb{R}^k\,, \tag{2.31}$$

where

$$(\mathbf{Z} \circ \mathbf{e}^{1/\gamma}(m(\Gamma A)^{1/k}) = \mathbf{Z}(m(\Gamma A)^{1/\gamma k}) \stackrel{\mathrm{d}}{=} m(\Gamma A)^{-1/k}\mathbf{Z}(1) \tag{2.32}$$

by Lemma 4.2.1.
(e) Note that

$$\begin{aligned} P(\alpha \in C(T_1(\epsilon)) &= P(Y(T_1(\epsilon)) - \alpha \in \phi(T_1(\epsilon))^{-1}\Gamma(T_1(\epsilon))A) \\ &= P(\Gamma(T_1(\epsilon))^{-1}\phi(T_1(\epsilon))[Y_1(T_1(\epsilon)) - \alpha] \in A, \det(\Gamma(T_1(\epsilon)) \neq 0) \\ &+ P(Y(T_1(\epsilon)) - \alpha \in \phi(T_1(\epsilon))^{-1}\Gamma(T_1(\epsilon))A; \det(\Gamma(T_1(\epsilon)) = 0) \end{aligned} \tag{2.33}$$

Since $T_1(\epsilon) \to \infty$ w.p.1 and $\Gamma(t) \to \Gamma$ w.p.1, where Γ is nonsingular, $P(\det(\Gamma(T_1(\epsilon))) = 0) \to 0$ as $\epsilon \downarrow 0$, so that the second term on the right in (2.33) is negligible. On the other hand, for the first term,

$$\begin{aligned} &\Gamma(T_1(\epsilon))^{-1}\phi(T_1(\epsilon))[Y(T_1(\epsilon)) - \alpha] = \Gamma(T_1(\epsilon))^{-1}\epsilon\phi(T_1(\epsilon))\epsilon^{-1}[Y(T_1(\epsilon) - \alpha] \\ &\Rightarrow \Gamma^{-1}m(\Gamma A)^{1/k}m(\Gamma A)^{-1/k}\Gamma\mathbf{Z}(1) = \mathbf{Z}(1) \end{aligned} \tag{2.34}$$

by parts (b) and (d). Hence, combining (2.33) and (2.34), we get

$$P(\alpha \in C(T_1(\epsilon))) \to P(\mathbf{Z}(1) \in A) = 1 - \delta\,, \tag{2.35}$$

because (2.4) holds with $L = \mathbf{Z}(1)$. ■

4.2.3. The Relative-Precision Sequential Estimator

We now state the analogous result for the relative-precision sequential estimator $T_2(\epsilon)$ in (2.11). Note that $T_2(\epsilon)$ behaves asymptotically like $T_1(\|\alpha\|\epsilon)$, as one would expect. In addition to the conditions in Theorem 4.2.2, we require that $Y(t) \to \alpha$ w.p.1 as $t \to \infty$. This is a reasonable condition, but it does not follow form the FCLT (2.14).

Theorem 4.2.2. *In addition to the conditions of Theorem 4.2.1, suppose that*

$$Y(t) \to \alpha \quad w.p.1 \quad in \quad \mathbb{R}^k \quad as \quad t \to \infty\,,$$

where $\|\alpha\| > 0$. *Then as* $t \to \infty$ *and* $\epsilon \to 0$

(a) $\phi(t)[m(C(t))^{1/k} + a(t)]/\|Y(t)\| \to \|\alpha\|^{-1} m(\Gamma A)^{1/k}$ *w.p.1,*

(b) $\epsilon\phi(T_2(\epsilon)) \to \|\alpha\|^{-1} m(\Gamma A)^{1/k}$ *w.p.1,*

(c) $\epsilon^{-1} m(C(T_2(\epsilon))^{1/k} \to \|\alpha\|$ *w.p.1,*

(d) $\epsilon^{-1}[Y(T_2(\epsilon)) - \alpha] \Rightarrow \|\alpha\| m(\Gamma A)^{-1/k}\Gamma Z(1)$ *in* $\mathbb{R}^k$

(e) $P(\alpha \in C(T_2(\epsilon))) \to 1 - \delta$ *(asymptotic validity).*

Since the proof of Theorem 4.2.2 closely parallels the proof of Theorem 4.2.1, we omit the proof of Theorem 4.2.2.

4.2.4. Analogs Based on a FWLLN

There are analogs of Theorems 4.2.1 and 4.2.2, where the SLLN for $\Gamma(t)$ in (2.15) is replaced by the weaker condition of a FWLLN. The w.p.1 limits in parts (a)–(c) of Theorems 4.2.1 and 4.2.2 are then replaced by FWLLNs and the CLT in (d) becomes a FCLT. Since the two results are similar, we only state the analog of Theorem 4.2.1.

Now we also generalize the framework by allowing a family of estimation processes indexed by ϵ. We start with processes $\{Y_\epsilon(t) : t \geq 0\}$ and $\{\Gamma_\epsilon(t) : t \geq 0\}$ for each $\epsilon > 0$. Then instead of (2.7), (2.9) and (2.13), let

$$\begin{aligned} C_\epsilon(t) &\equiv Y_\epsilon(t) - \phi(t)\Gamma_\epsilon(t)A\,, \\ T_{1\epsilon} &\equiv \inf\{t \geq 0 : m(C_\epsilon(t))^{1/k} + a(t) < \epsilon\}\,, \\ \mathbf{Y}_\epsilon(t) &= \phi(\epsilon^{-1})[Y_\epsilon(t/\epsilon) - \alpha], \quad t \geq 0\,. \end{aligned} \tag{2.36}$$

Then the limit will be for the following processes: For that purpose, we define the following random elements of D:

$$\begin{aligned} \mathbf{\Gamma}_\epsilon(t) &\equiv \Gamma_\epsilon(t/\epsilon), \quad t > 0\,, \\ \mathbf{U}^1_\epsilon(t) &\equiv \phi(\epsilon^{-1})m(C_\epsilon(t/\epsilon))^{1/k}\,, \\ \mathbf{U}^2_\epsilon(t) &\equiv \epsilon T_{1\epsilon}(1/t\phi(\epsilon^{-1}))\,, \\ \mathbf{U}^3_\epsilon(t) &\equiv \epsilon^{-1}m(C_\epsilon(T_{1\epsilon}(\epsilon/t)))^{1/k}\,, \\ \mathbf{Z}_\epsilon(t) &\equiv \epsilon^{-1}[Y_\epsilon(T_{1\epsilon}(\epsilon/t)) - \alpha]\,. \end{aligned} \tag{2.37}$$

Theorem 4.2.3. *Let the topology on D be one of WM_2, SM_2, WM_1, SM_1, WJ_1 or SJ_1 throughout. Suppose that the assumptions of Theorem 4.2.1 hold, except that (2.13) is replaced by (2.36) and condition (2.15) is replaced by*

$$\mathbf{\Gamma}_\epsilon \Rightarrow \Gamma\mathbf{1} \quad in \quad D^{k^2} \quad as \quad \epsilon \downarrow 0\,, \tag{2.38}$$

for Γ_ϵ in (2.37) and $\mathbf{1}(t) = (1, \ldots, 1)$ for all $t > 0$. Then

$$(\mathbf{\Gamma}_\epsilon, \mathbf{U}^1_\epsilon, \mathbf{U}^2_\epsilon, \mathbf{U}^3_\epsilon, \mathbf{Z}_\epsilon) \Rightarrow (\Gamma\mathbf{1}, \mathbf{U}^1, \mathbf{U}^2, \mathbf{U}^3, \mathbf{Z}^1) \quad in \quad D^{4+k} \quad as \quad \epsilon \downarrow 0\,, \tag{2.39}$$

for $(\mathbf{U}^1_\epsilon, \mathbf{U}^2_\epsilon, \mathbf{U}^3_\epsilon, \mathbf{Z}_\epsilon)$ in (2.37), where

$$\begin{aligned} \mathbf{U}^1(t) &\equiv t^{-\gamma} m(\Gamma A)^{1/k}, \quad \mathbf{U}^2(t) \equiv t^{1/\gamma}(\Gamma A)^{1/\gamma k} \\ \mathbf{U}^3(t) &= t^{-1} \quad and \quad \mathbf{Z}'(t) \equiv m(\Gamma A)^{-1/k}\Gamma\mathbf{Z}(t^{1/\gamma})\,. \end{aligned} \tag{2.40}$$

Moreover,

$$P(\alpha \in C_\epsilon(T_{1\epsilon}(\epsilon))) \to 1 - \delta \text{ (asymptotic validity)}\,. \tag{2.41}$$

In preparation for the proof of Theorem 4.2.3, we prove a lemma.

Lemma 4.2.2. *If $x_i \in D([a,b], \mathbb{R})$ for $i = 1, 2$, where $x_1(t)$, $x_2(t) \geq c > 0$ for all t, then*

$$\|y_1 - y_2\| \leq c^{-2}\|x_1 - x_2\|$$

for $y_i(t) = 1/x_i(t)$, $a \leq t \leq b$.

Proof. Note that

$$|y_1(t) - y_2(t)| = \frac{|x_2(t) - x_1(t)|}{|x_1(t)| \cdot |x_2(t)|} \leq c^{-2}|x_2(t) - x_1(t)|\,. \quad \blacksquare$$

Corollary 4.2.1. *If $x_i \in D([a,b], \mathbb{R})$, $x_i(t) \geq c > 0$, and $y_i(t) = 1/x_i(t)$, $a \leq t \leq b$, $i = 1, 2$ then*

$$d(y_1, y_2) \leq (c^{-2} \vee 1) d(x_1, x_2)$$

where d is one of the J_1, M_1 or M_2 metrics.

Proof. To illustrate, we do the J_1 case:

$$\begin{aligned} d(y_1, y_2) &= \inf_{\lambda \in \Lambda}\{\|y_1 - y_2 \circ \lambda\| \vee \|\lambda - e\|\} \\ &\leq \inf_{\lambda \in \Lambda}\{c^{-2}\|x_1 - x_2 \circ \lambda\| \vee \|\lambda - e\|\} \\ &\leq (c^{-2} \vee 1) d(x_1, x_2)\,. \quad \blacksquare \end{aligned}$$

Proof of Theorem 4.2.3. First since the limit $\Gamma\mathbf{1}$ in (2.38) is deterministic, the two limits in (2.38) and (2.14) hold jointly (where $\mathbf{Y}_\epsilon$ is defined by (2.36) instead of (2.13)), by virtue of Theorem 11.4.5 in the book. Given those limits, we can apply the Skorohod representation theorem, as $\epsilon \downarrow 0$ through an arbitrary sequence, to replace the convergence in distribution by special versions converging w.p.1. Let the special versions be represented by the same notation. Since $\mathbf{1} \in C$, the convergence $\mathbf{\Gamma}_\epsilon \to \Gamma\mathbf{1}$ in $D((0,\infty), \mathbb{R}^{d^2})$ is equivalent to uniform convergence over bounded intervals. Then, as in the proof of Theorem 4.2.1 (a), apply the bounded convergence theorem to get $m(\Gamma_\epsilon(t/\epsilon)A)^{1/k} \to m(\Gamma A)^{1/k}$ w.p.1 uniformly for $t \in [t_0, t_1]$ for any t_0, t_1 with $0 < t_0 < t_1 < \infty$. This yields w.p.1 convergence in $D((0,\infty), \mathbb{R})$ for the special versions. Since $\phi(t/\epsilon)$ a $(t/\epsilon) \to 0$ as $\epsilon \downarrow 0$ uniformly in t for $t > t_0$, we obtain

$$\phi(t/\epsilon)V_\epsilon(t/\epsilon) \to m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.42}$$

uniformly in $[t_0, t_1]$ for the special versions. Since $a(t/\epsilon) = o(\phi(t/\epsilon)^{-1})$, (2.42) implies that

$$\phi(t/\epsilon)m(C_\epsilon(t/\epsilon))^{1/k} \to m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \to 0 \tag{2.43}$$

uniformly in $[t_0, t_1]$. However, since $\phi \in \mathcal{R}(\gamma)$, $\phi(\epsilon^{-1})/\phi(t/\epsilon) \to t^{-\gamma}$ as $\epsilon \downarrow 0$ uniformly on $[t_0, t_1]$, by Theorem A.5 in Appendix A of the book. Thus,

$$\phi(\epsilon^{-1})V_\epsilon(t/\epsilon) \to t^{-\gamma}m(\Gamma A)^{1/k} \tag{2.44}$$

and

$$\phi(\epsilon^{-1})m(C_\epsilon(t/\epsilon))^{1/k} \to t^{-\gamma}m(\Gamma A)^{1/k} \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.45}$$

uniformly in $[t_0, t_1]$, again for the special versions, which implies the FCLT conclusion for $\mathbf{U}_\epsilon^1$ in $D((0,\infty), \mathbb{R})$. Turning to $\mathbf{U}_\epsilon^2$, we will show for the special versions that

$$\begin{aligned}
\epsilon T_{1\epsilon}(1/t\phi(\epsilon^{-1})) &= \inf\{s \geq 0 : \phi(\epsilon^{-1})V_\epsilon(s/\epsilon) < t^{-1}\} \\
&= \inf\{s \geq 0 : \phi(\epsilon^{-1})^{-1}V_\epsilon(s/\epsilon)^{-1} > t\} \\
&\to \inf\{s \geq 0 : s^\gamma m(\Gamma A)^{-1/k} > t\} \\
&= t^{1/\gamma}m(\Gamma A)^{1/\gamma k} \ .
\end{aligned} \tag{2.46}$$

uniformly in $[t_0, t_1]$. In the first line of (2.46), without loss of generality, we can replace $\phi(\epsilon^{-1})V_\epsilon(s/\epsilon)$ by $\max\{\phi(\epsilon^{-1})V_\epsilon(s/\epsilon), (2t_1)^{-1}\}$. Then we can invoke Corollary 4.2.1 above to show that the third line follows from (2.44). For $\mathbf{U}_\epsilon^3$, apply the continuous-mapping theorem with the composition map, using Theorem 13.2.3 in the book and (2.43) and (2.46) here, to get

$$\phi(\epsilon^{-1})m(C_\epsilon(T_{1\epsilon}(1/t\phi(\epsilon^{-1}))))^{1/k} \to t^{-1} \quad \text{as} \quad \epsilon \to 0 \tag{2.47}$$

uniformly in $[t_0, t_1]$ or, equivalently,

$$\epsilon^{-1} m(C_\epsilon(T_{1\epsilon}(\epsilon/t)))^{1/k} \to t^{-1} \quad \text{as} \quad \epsilon \to 0 \tag{2.48}$$

uniformly in $[t_0, t_1]$. Next, for $\mathbf{Z}_\epsilon$, apply the composition map again with the FCLT in (2.14) and the limit for $\mathbf{U}_\epsilon^2$ in (2.48) and part (b) to get

$$\mathbf{Z}_{\phi(\epsilon^{-1})^{-1}} \to \mathbf{Z}' \quad \text{in} \quad D((0,\infty), \mathbb{R}^k) \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.49}$$

where

$$\mathbf{Z}'(t) \equiv \Gamma \mathbf{Z}(t^{1/\gamma} m(\Gamma A)^{1/\gamma k}) \stackrel{\mathrm{d}}{=} m(\Gamma A)^{-1/k} \Gamma \mathbf{Z}(t^{1/\gamma})$$

and the topology is the same as for (2.14). Clearly, $\mathbf{Z}_\epsilon \to \mathbf{Z}'$ in $D((0,\infty), \mathbb{R}^k)$ as well. Finally, for (2.41), apply the projection map for $t = 1$ with the result $\mathbf{Z}_\epsilon \to \mathbf{Z}'$ just established. Then use the argument for Theorem 4.2.1 (e). ■

4.2.5. Examples

We conclude this section by giving several examples. We illustrate how the theorems can be applied by discussing a few specific estimation settings. These examples show that FCLT requirement for the estimation process Y in (2.14) is a mild hypothesis that is satisfied in virtually all practical contexts. However, some work may be required to establish the SLLN or FWLLN for the estimators $\Gamma(t)$ of the scaling matrix Γ. Our last example shows that we cannot instead use weak consistency of $\Gamma(t)$.

Example 4.2.1. (*Sample mean of IID random variables*). Suppose that α can be represented as $\alpha = EX$ for some real-valued r.v. X. For example, α might correspond to the expected number of customers served in a queue over the time interval $[0, T]$. Then α can be estimated by generating i.i.d. replicates X_1, $X_2, \ldots$ of the r.v. X; the resulting estimator for α is then the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. The corresponding estimation process is $Y(t) = \bar{X}_{\lfloor t \rfloor}$, where $\lfloor t \rfloor$ is the greatest integer less than t and $\bar{X}_0 = 0$. If $EX^2 < \infty$, then Donsker's theorem, Theorem 4.3.2 in the book, asserts that the FCLT in (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$ in (2.13), $\Gamma = \sigma$, where $\sigma^2 = \operatorname{var} X$, and $\mathbf{Z}(t) = \mathbf{B}(t)/t$, where $\mathbf{B}$ is Brownian motion. Note that $\mathbf{Z}(1) =_d N(0,1)$. The typical choice for the set A in this setting is the interval $[-z(\delta),\, z(\delta)]$, where $z(\delta)$ is chosen to satisfy $P(N(0,1) \le z(\delta)\} = 1 - \delta/2$. Of course, it is well known that

$$\Gamma_n \equiv \left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} \to \sigma \quad \text{w.p.1 as} \quad n \to \infty \,. \tag{2.50}$$

Suppose that $\sigma^2 > 0$. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we have the strong consistency required by Theorems 4.2.1 and 4.2.2. Hence both the absolute-precision and relative-precision stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid for this example when the precision-constant ϵ shrinks to 0. In this setting, Theorems 4.2.1 and 4.2.2 reproduce the classical results of Chow and Robbins (1965), Starr (1966) and Nadas (1969); see Chapter 7 of Siegmund (1985) and Section 8.8 of Wetherill and Glazebrook (1986). (See Anscombe (1952, 1953) for related earlier work.) Implementation considerations are discussed in Law, Kelton and Koenig (1981).

Example 4.2.2. (*The sample mean of IID random vectors*). Now we consider the case in which α can be represented as $\alpha = EX$, where X is $\mathbb{R}^k$-valued. Assume that $E\|X\|^2 < \infty$. As in Example 4.2.1, we can estimate α via the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, where X_i's are i.i.d. copies of X. Setting $Y(t) = \bar{X}_{\lfloor t \rfloor}$, we obtain the FCLT (2.14) from the k-dimensional version of Donsker's theorem, Theorem 4.3.5 in the book, where now $\mathbf{Z}(t) = \mathbf{B}(t)/t$, $\mathbf{B}$ is k-dimensional standard Brownian motion (composed of k independent one-dimensional standard Brownian motions) and $\Gamma\Gamma^t$ is the covariance matrix C of X. We assume that C is positive definite. Note that $\mathbf{Z}(1) = \mathbf{B}(1) =_d N(0, I)$, where I is the identity matrix. In this k-dimensional setting, we can assume that A is the k-sphere $\{x : \|x\| \le w(\delta)\}$, where $w(\delta)$ is chosen so that

$$P\{\|N(0,I)\|^2 \le w^2(\delta)\} = P\{\mathcal{X}_k^2 \le w^2(\delta)\} = 1 - \delta \ , \tag{2.51}$$

with $\mathcal{X}_k^2$ being a chi-squared r.v. with k degrees of freedom. Let

$$C_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^t - \bar{X}_n \bar{X}_n^t \tag{2.52}$$

(writing all k-vectors as column vectors). Then $C_n \to C$ a.s. as $n \to \infty$. Let Γ_n be obtained by taking the Cholesky factorization of C_n, so that Γ_n is a lower triangular matrix such that $C_n = \Gamma_n \Gamma_n^t$; see pages 164 and 165 of Bratley, Fox and Schrage (1987). It follows that $\Gamma_n \to \Gamma$ w.p.1 as $n \to \infty$, since Cholesky factors are continuous at positive definite matrices. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we again have the strong consistency required by Thoerems 4.2.1 and 4.2.2. Thus we have proved that the absolute-precision and relative-precision stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid for sequential stopping of multiple performance measure stochastic simulations. In this setting, Theorems 4.2.1 and 4.2.2 reproduce results by Gleser (1965), Albert (1966) and Srivastava (1967); see Section 5.5 of Govindarajulu (1987).

Example 4.2.3. (*Functions of sample means*). Let X be an $\mathbb{R}^k$-valued random vector and let $\mu = EX$. Suppose that α can be represented as $\alpha = g(\mu)$ for some (known) real-valued function $g : \mathbb{R}^k \to \mathbb{R}$. An example of this occurs in the ratio estimation setting, in which $k = 2$ and $g(x, y) = x/y$. Because the steady state of a regenerative stochastic process can be expressed as a ratio of two means, this estimation setting subsumes that of regenerative steady-state simulation. Of course, this observation lies at the heart of the regenerative method of steady-state simulation; see, for example, Crane and Lemoine (1977).

In this nonlinear setting, we estimate α via $Y(t) = g(\bar{X}_{\lfloor t \rfloor})$, where X_i are i.i.d. random vectors as in Example 4.2.2. Suppose that $E\|X\|^2 < \infty$ and that g is continuously differentiable in a neighborhood of μ. In addition, we require that $\nabla g(\mu) \neq 0$ and that the covariance matrix C of X is positive definite. Then Theorem 3 of Glynn and Whitt (1992b) implies that the FCLT in (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$, $\mathbf{Z}(t) = \mathbf{B}(t)/t$ and $\Gamma = \sigma$ as in Example 4.2.1, but with

$$\sigma = (\nabla g(\mu)^t C \nabla g(\mu))^{1/2} \ .$$

Let C_n be defined as in Example 4.2.2 and note that

$$[\nabla g(Y(t))^t C_{\lfloor t \rfloor} \nabla g(Y(t))]^{1/2} \to \sigma \text{ w.p.1 as } t \to \infty \, .$$

Hence we have the strong consistency required for the application of Thoerems 4.2.1 and 4.2.2. As a consequence, we are assured that the stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are again asymptotically valid in this estimation setting. In particular, in the regenerative simulation setting, we recover the asymptotic theory developed by Lavenberg and Sauer (1977).

Example 4.2.4. (*The jackknife*). Consider the estimation problem of Example 4.2.3 in which our goal is to estimate $\alpha = g(\mu)$, where μ can be expressed as $\mu = EX$ and g is real-valued. One practical difficulty with the estimator suggested in Example 4.2.3 is that it tends to be significantly affected by bias problems induced by the presence of the nonlinearity in g. One way to address the small-sample bias problem that this nonlinearity creates is to jackknife the estimator. Specifically, let $\alpha(n) = g(\bar{X}_n)$ and, for $1 \leq i \leq n$, let

$$\begin{aligned} \bar{X}_{in} &= \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} X_j, \quad \alpha_i(n) = g(\bar{X}_{in}) \ , \\ \tilde{\alpha}_i(n) &= n\alpha(n) - (n-1)\alpha_i(n) \ . \end{aligned} \tag{2.53}$$

Then the estimator $Y_n = n^{-1}\sum_{i=1}^n \tilde{\alpha}_i(n)$ is the *jackknife estimator* of α. Let $Y(t) = Y_{\lfloor t\rfloor}$. It is shown in Glynn and Heidelberger (1989) that if $E\|X\|^3 < \infty$ and g is twice continuously differentiable in a neighborhood of μ, then the FCLT in (2.14) holds where σ and $Z(t)$ are as in Example 4.2.3. Since the form of the FCLT is the same as for Example 4.2.3, the jackknife has the same asymptotic efficiency as the estimator of Example 4.2.3. However, as argued in Miller (1964, 1974), the jackknife estimator typically possesses superior small-sample bias properties.

Two estimators for the scaling constant $\sigma = [\nabla g(\mu)^t C \nabla g(\mu)]^{1/2}$ are possible. One approach is to use the estimator $\sigma(t) = [\nabla g(Y(t))^t C_{\lfloor t\rfloor} \nabla g(Y(t))]^{1/2}$ suggested in Example 4.2.3. Theorem 4(i) of Glynn and Heidelberger (1989) shows that $Y(t) \to \alpha$ w.p.1 as $t \to \infty$, under the conditions stated here. Since $C_n \to C$ w.p.1, it follows that $\sigma(t) \to \sigma$ w.p.1 as $t \to \infty$. Hence sequential stopping procedures based on the jackknife point estimator and the "variance" estimator $\sigma^2(t)$ are asymptotically valid by Theorems 4.2.1 and 4.2.2, provided that $\sigma^2 > 0$.

An alternative estimator for the scaling constant σ is given by the jackknife variance estimator $\sigma_J(t)$:

$$\sigma_J(t) = \left(\frac{1}{\lfloor t\rfloor} \sum_{i=1}^{\lfloor t\rfloor} (\tilde{\alpha}_i(\lfloor t\rfloor) - Y(t))^2 \right)^{1/2} . \tag{2.54}$$

Although it is known that $\sigma_J^2(t) \Rightarrow \sigma^2$ as $t \to \infty$ under suitable regularity conditions, we need convergence w.p.1 in order to satisfy the hypothesis of Theorems 4.2.1 and 4.2.3. However, Theorem 4 of Glynn and Whitt (1992a) establishes the following result.

Theorem 4.2.4. *If g is continuously differentiable in a neighborhood of μ and $E|X|^2 < \infty$, then*

$$\sigma_J^2(t) \to \sigma^2 = \nabla g(\mu)^t C\ \nabla g(\mu) \ \textit{w.p.1 as } t \to \infty \tag{2.55}$$

for $\sigma_J^2(t)$ in (2.54). Thus the sequential stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ may be applied to jackknife point estimators in conjunction with the jackknifed variance estimator $\sigma_J^2(t)$.

Example 4.2.5. (*A steady-state mean*). Suppose that our goal is to estimate the steady-state mean vector α of an $\mathbb{R}^k$-valued stochastic process $X = \{X(t) : t \geq 0\}$. We assume that X satisfies an FCLT, namely,

$$\mathbf{X}_\epsilon \Rightarrow \Gamma\mathbf{B} \quad \text{in} \quad D((0,\infty), \mathbb{R}^k) \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.56}$$

where

$$\mathbf{X}_\epsilon(t) \equiv \epsilon^{-1}\left(\int_0^{t/\epsilon} X(s)ds - t\alpha\right), \quad t > 0. \tag{2.57}$$

and $\mathbf{B}$ is a standard $\mathbb{R}^k$-valued Brownian motion. It is easily shown that (2.56) implies that

$$Y(t) \equiv t^{-1}\int_0^t X(s)ds \Rightarrow \alpha \quad \text{as} \quad t \to \infty. \tag{2.58}$$

Hence (2.56) implies that the centering vector α appearing there is indeed the steady-state mean of X. Another easy consequence of (2.56) is that the FCLT (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$ and $\mathbf{Z}(t) = \mathbf{B}(t)/t$.

It turns out that (2.56) is typically satisfied for most "real-world" steady-state simulations. In particular, a great variety of different assumptions on the structure of the process X give rise to FCLTs of the form (2.56); see Section 4.4 in the book and Section 2.3 here.

The primary difficulty in applying Theorems 4.2.1–4.2.3 arises in the construction of a process $\Gamma(t)$ such that $\Gamma(t) \to \Gamma$ w.p.1 as $t \to \infty$ or $\mathbf{\Gamma}_\epsilon \Rightarrow \Gamma\mathbf{1}$ in $D(0,\infty)$ as $\epsilon \downarrow 0$. Since $\Gamma\Gamma^t$ is the covariance matrix of the limiting Brownian motion, this is equivalent to the construction of a strongly consistent estimator $C(t)$ for the *time-average covariance matrix* $C = \Gamma\Gamma^t$ of X. In general, this is known to be a challenging problem.

Suppose that X is regenerative, with regeneration times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$. Suppose that $E(\int_{\tau_1}^{\tau_2} |X(s) - \alpha(ds)^2 < \infty$ and that $E(\tau_2 - \tau_1) < \infty$. Let $N(t) = \max\{n \geq 0 : \tau_n \leq t\}$. Then it is easily proved that

$$C(t) = \frac{1}{t}\sum_{i=1}^{N(t)} \int_{\tau_i - 1}^{\tau_i} [X(s) - Y(t)][X(s) - Y(t)]^t ds \tag{2.59}$$

is strongly consistent for C, where $C = \Gamma\Gamma^t$ and Γ is the scaling matrix appearing in (2.56). Thus when X is regenerative, the sequential stopping rules $T_1(\epsilon)$ and $T_2(\epsilon)$ are asymptotically valid. Of course, when X is scalar, we already established this result in Example 4.2.3.

For nonregenerative processes, less is known about the strong consistency of estimators $C(t)$ for the steady-state covariance matrix. However, Glynn and Iglehart (1988) and Damerdji (1991, 1994) have recently used strong approximation techniques to establish strong consistency for a broad class of estimators for C. Thus Theorems 4.2.1 and 4.2.2 prove that these estimators do indeed lead to asymptotically valid sequential procedures.

Our theory for this example provides theoretical support complementing previous work by Fishman (1977), Law and Carson (1979) and Law and Kelton (1982). They develop specific empirically based sequential stopping rules for steady-state simulations.

Example 4.2.6. (*Kiefer-Wolfowitz stochastic approximation*). This example is interesting, in part, because it illustrates that the FCLT (2.14) can hold for the estimator with a subcanonical convergence rate; in particular, here $\phi(\epsilon^{-1}) = \epsilon^{-1/3}$. For other examples of noncanonical estimator convergence rates, see Fox and Glynn (1989) and Sections 5 and 6 of Glynn and Whitt (1992b). Suppose that we are given a real-valued smooth function $\beta(\theta)$, which can be represented as $\beta(\theta) = EZ(\theta)$. Assume that our goal is to compute the parameter $\alpha \equiv \theta^*$ minimizing β. If θ is scalar, we can apply the following Kiefer-Wolfowitz stochastic approximation algorithm:

$$\theta_{n+1} = \theta_n - c_n X_{n+1} \ , \tag{2.60}$$

where $\{c_n : n \geq 0\}$ is a sequence of (deterministic) nonnegative constants,

$$\begin{aligned} &P(X_{n+1} \in A | \theta_0, X_0, \ldots, \theta_n, X_n) = \\ &\qquad P([Z(\theta_0 + h_{n+1}) - Z(\theta_0 - h_{n+1}]/2h_{n+1} \in A) \, , \end{aligned} \tag{2.61}$$

$Z(\theta_0 + h_{n+1})$ and $Z(\theta_0 - h_{n+1})$ are generated independently of one another and $\{h_n : n \geq 1\}$ is another sequence of deterministic constants. Suppose that $c_n = cn^{-1}$ and $h_n = hn^{-1/3}$, c, $h > 0$. Let $Y(t) = \theta_{\lfloor t \rfloor}$. For this problem, Ruppert (1982) showed that under suitable regularity conditions, the FCLT in (2.14) holds for $\mathbf{Y}_\epsilon$ in (2.13) with $\phi(\epsilon^{-1}) = \epsilon^{-1/3}$, $\Gamma = \kappa$, $\mathbf{Z}(t) = t^{-b}\mathbf{B}(t^{2\eta+1})$, $\mathbf{B}$ is a standard Brownian motion, $b = c\beta''(\theta^*)$, $\eta = b - 5/6$, $\kappa^2 = c^2\sigma^2/(2\eta+1)(4h^2)$ and $\sigma^2 = 2var\ \mathbf{Z}(\theta^*)$.

The construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ involves more work. For some directions on how to obtain such an estimator, see page 189 of Venter (1967).

Example 4.2.7. (*Robbins-Monro stochastic approximation*). As in Example 4.2.6, suppose that our goal is to estimate the minimizer θ^* of a smooth function β: $\mathbb{R} \to \mathbb{R}$. However, we assume here that we can represent the derivative β' as an expectation; that is, there exists a process $Z(\theta)$ such that $\beta'(\theta) = EZ(\theta)$. [In Example 4.2.6 we assumed only that the function values $\beta(\theta)$ could be represented as expectations.] To calculate θ^* in this setting, we can use the Robbins-Monro stochastic approximation algorithm, which is based on (2.60), where $\{c_n : n \geq 0\}$ is sequence of (deterministic) nonnegative constants and

$$P(X_{n+1} \in A | \theta_0, X_0, \ldots, \theta_0, X_n) = P(Z(\theta_n) \in A) . \tag{2.62}$$

Suppose that our estimator is $Y(t) = \theta_{\lfloor t \rfloor}$ and $c_n = cn^{-1}$ with $c > 0$. Then Ruppert (1982) showed that under suitable regularity hypotheses, the FCLT in (2.14) holds for $\mathbf{Y}_\epsilon$ in (2.13) with $\phi(\epsilon^{-1}) = \epsilon^{-1/2}$, $\Gamma = \kappa$, $\mathbf{Z}(t) = t^{-(D+1)}\mathbf{B}(t^{2D+1})$, $D = c\beta'(\theta^*) - 1$, $\kappa^2 = c^2\sigma^2(2D+1)^{-1}$, $\sigma^2 = var\ \mathbf{Z}(\theta^*)$ and $\mathbf{B}$ is a standard Brownian motion.

Construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ follows from results established by Venter (1967). When this estimator is used, the sequential stopping rule $T_1(\epsilon)$ reduces to one studied by McLeish (1976).

Example 4.2.8. (*The Hill estimator*). The framework of Theorems 4.2.1–4.2.3 has been made quite general, so that there can be many applications. One intended application is to estimation problems associated with heavy-tailed probability distributions and long-range dependence. If we use the direct (naive) estimators, e.g., the time average for the steady-state mean, then we anticipate that the FCLT in (2.14) will typically hold with ϕ in (2.13) satisfying $\phi(\epsilon^{-1})/\epsilon^{-1/2} \to 0$ as $\epsilon \downarrow 0$. A common case would by $\phi(\epsilon^{-1}) = \epsilon^{-\gamma}$ or $\phi \in \mathcal{R}(\gamma)$ for $0 < \gamma < 1/2$. A major new difficulty, however, is that now the scaling exponent γ is typically unknown.

Thus, attention naturally shifts to estimating the scaling parameter γ. Estimating the parameter γ is challenging even from observations of i.i.d. random variables. One approach is via the Hill estimator. Recent results of Resnick and Stărică (1997) show that Theorems 4.2.1–4.2.3 can be applied.

The setting is a sequence $\{X_n : n \geq 1\}$ of i.i.d. positive random variables hoaving cdf F, where $F^c \equiv 1 - F \in \mathcal{R}(-\alpha)$ for $\alpha > 0$, i.e.

$$F^c(tx)/F^c(t) \to x^{-\alpha} \quad \text{as} \quad t \to \infty . \tag{2.63}$$

The goal is to estimate the tail index α. For n given, let $X_{(i)}$ be the i^{th} largest among the first n. The Hill estimator based on the k upper order statistics is

$$H_{k,n} = k^{-1} \sum_{i=1}^{k} \log \left(\frac{X_{(i)}}{X_{(k+1)}} \right) . \tag{2.64}$$

The Hill estimator is known to be consistent if $k \equiv k(n)$ satisfies $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$. Given a specific function $k(n)$, the Hill estimator is a single sequence of random variables $\{H_{k(n),n} : n \geq 1\}$. Resnick and

Stărică (1997) show that the Hill estimator also satisfies a FCLT. To state it, let

$$\mathbf{Y}_n(t) \equiv H_{\lceil k_n t \rceil, n}, \quad t \geq 0 , \tag{2.65}$$

where $\lceil x \rceil$ is the least integer greater than or equal to x. The FCLT states that, under regularity conditions, including $k(n) \to \infty$ and $k(n)/n \to 0$,

$$k(n)[\mathbf{Y}_n - \alpha^{-1}\mathbf{1}] \Rightarrow \alpha^{-1}\mathbf{Z} \quad \text{in} \quad D \quad \text{as} \quad n \to \infty , \tag{2.66}$$

with $\mathbf{Z}(t) = t^{-1}\mathbf{B}(t)$, where $\mathbf{B}$ is standard Brownian motion. Notice that here we use the more general framework in (2.36) in which there is a family of estimation processes indexed by $\epsilon > 0$. (Here we have used $n \to \infty$ instead of $\epsilon \downarrow 0$.)

Also notice that in this special case the scaling matrix Γ in (2.14) is just α^{-1}. So, with $\mathbf{Y}_n$ in (2.65), we estimate α and Γ simultaneously. As a consequence of the FCLT in (2.66), we have the associated FWLLN

$$\mathbf{Y}_n \Rightarrow \alpha^{-1}\mathbf{1} \quad \text{in} \quad D \quad \text{as} \quad n \to \infty \tag{2.67}$$

needed in Theorem 4.2.3. It is also known that $Y_n(t) \to \alpha^{-1}$ w.p.1 as $n \to \infty$ under regularity conditions.

Given the FCLT in (2.66) and the FWLLN in (2.67), the conditions of Theorem 4.2.3 are satisfied. Hence sequential stopping rules are asymptotically valid for the Hill estimator too.

Example 4.2.9. (*Sample mean with infinite variance*). One can also estimate a mean by the sample mean of i.i.d. random variables when the random variables X_i have finite mean but infinite variance. As in Example 4.2.1, the estimation process can be $Y(t) = \bar{X}_{\lfloor t \rfloor}$, where $\bar{X}_0 = 0$, although it is often better to use alternative robust estimators such as trimmed means or to estimate other quantities such as the median. Under regularity conditions, FCLT (2.14) is valid with $\phi \in \mathcal{R}(1 - \alpha^{-1})$ for some α, $1 < \alpha < 2$, where ϕ is the scaling function in (2.13). The topology on D can be the J_1 topology. The limit process $\mathbf{Z}(t)$ is then $t^{-1}\mathbf{S}_\alpha(t)$, where $\{\mathbf{S}_\alpha(t) : t \geq 0\}$ is a stable process of index α, which depends on two parameters in addition to α: a scale parameter σ and a skewness parameter β, $-1 \leq \beta \leq 1$. Unfortunately, in order to form confidence sets we need to estimate the scaling function ϕ and the parameters σ and β.

Suppose that we consider the special case in which X_i is nonnegative and is assumed to have an asymptotic power tail, i.e.

$$F^c(t) \equiv P(X > t) \sim At^{-\alpha} \quad \text{as} \quad t \to \infty \tag{2.68}$$

for positive constants A and α, $1 < \alpha < 2$. Under condition (2.68), the FCLT (2.14) holds with $\phi(\epsilon^{-1}) = \epsilon^{-(1-\alpha^{-1})}$ and limit process $\Gamma Z(t)$ where $Z(t)$ is a stable process with index α, scale $\sigma = 1$ and skewness 1. Hence, in this special case it suffices to estimate only the two parameters α and Γ.

Suppose that $\hat{\alpha}_\epsilon$ is an estimate of α with the property that

$$(\hat{\alpha}_\epsilon^{-1} - \alpha^{-1}) \log(\epsilon^{-1}) \to 0 \quad \text{w.p.1} \quad \text{as} \quad \epsilon \downarrow 0 \,. \tag{2.69}$$

Given (2.69),

$$\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1}) \equiv \epsilon^{-(1-\hat{\alpha}_\epsilon^{-1})}/\epsilon^{-(1-\alpha)} \,, \tag{2.70}$$

and

$$\log[\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1})] = (\alpha^{-1} - \hat{\alpha}_\epsilon^{-1}) \log(\epsilon^{-1}) \to 0 \quad \text{w.p.1} \quad \text{as} \quad \epsilon \downarrow 0 \,, \tag{2.71}$$

so that

$$\hat{\phi}(\epsilon^{-1})/\phi(\epsilon^{-1}) \to 1 \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.72}$$

and the FCLT (2.14) holds with the estimator $\hat{\phi}(\epsilon^{-1}) \equiv \epsilon^{-(1-\hat{\alpha}_\epsilon^{-1})}$ used in place of the scaling function $\phi(\epsilon^{-1}) = \epsilon^{-(1-\alpha^{-1})}$. Hence it only remains to estimate the scale parameter Γ. Given that (2.68) holds, the scale parameter is

$$\Gamma = (A/A_\alpha)^{1/\alpha} \tag{2.73}$$

for A in (2.68) and

$$A_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} \,. \tag{2.74}$$

Hence it suffices to estimate the asymptotic constant A in (2.68). We can estimate in various ways if we estimate the cdf in (2.68) by the empirical cdf.

Hence, under regularity conditions, the sequential stopping rules will again be asymptotically valid. However, in this situation it is often much better to use different (robust) estimators for the mean or to estimate different quantities, such as the median or other percentiles.

Example 4.2.10. (*A counterexample for weak consistency*). Since the SLLN or FWLLN for $\Gamma(t)$ is relatively difficult to establish, it is natural to ask if the weak consistency $\Gamma(t) \Rightarrow \Gamma$ as $t \to \infty$ in (2.6) might not be enough to ensure asymptotic validity of the sequential stopping rules.

Unfortunately, however, weak consistency of $\Gamma(t)$ is not enough. The difficulty is in establishing the in-probability analog of Theorem 4.2.1 (b).

We now give a direct counterexample. Consider Example 4.2.1 and the process $\Gamma(t)$ defined there. Let N be a unit rate Poisson process independent of $\{X_i : i \geq 1\}$ and let $T_1, T_2, \ldots$ be the jump times of the process N. Suppose that

$$\tilde{\Gamma}_{(t)} = \begin{cases} \Gamma(t), & t \notin \cup_{n=1}^{\infty}[T_n, T_n + 1/n), \\ \\ 0, & t \in \cup_{n=1}^{\infty}[T_n, T_n + 1/n). \end{cases} \tag{2.75}$$

Then

$$\begin{aligned} P(\tilde{\Gamma}(t) \neq \Gamma(t)) &= P\left(t \in \left[T_{N(t)}, T_{N(t)} + \frac{1}{N(t)}\right]\right) \\ &\leq P(t - T_{N(t)} \leq \epsilon) + P\left(N(t) \leq \frac{1}{\epsilon}\right) \end{aligned} \tag{2.76}$$

for ϵ arbitrary. Letting $t \to \infty$, we find that $\limsup_{t\to\infty} P(\tilde{\Gamma}(t) \leq \Gamma(t)) = 1 - \exp(-\epsilon)$ (recall that the equilibrium age distribution of N is exponential with mean 1). Since ϵ was arbitrary, it follows that $P(\tilde{\Gamma}(t) \neq \Gamma(t)) \to 0$ as $t \to \infty$. Then it is evident that $\tilde{\Gamma}(t) \Rightarrow \sigma$ as $t \to \infty$, since $\Gamma(t) \to \sigma$ w.p.1 as $t \to \infty$.

Now, in the setting of Example 4.2.1 using $\tilde{\Gamma}(t)$,

$$\tilde{T}_1(\epsilon) = \inf\left\{t \geq 0 : z(\delta)\left(\frac{\Gamma(t)}{\sqrt{t}} + a(t)\right) \leq \epsilon\right\} . \tag{2.77}$$

Put $a(t) = 1/t$. Then clearly $z(\delta)(\tilde{\Gamma}(s)/\sqrt{s} + 1/s) \geq z(\delta)/t$ and $s \leq t$, so $\tilde{T}_1(z(\delta)/t) \geq t$. On the other hand, $\tilde{\Gamma}(T_{N(t)+1}) = 0$, so $\tilde{T}_1(z(\delta)/t) \leq T_{N(t)+1}$. By the SLLN, $t^{-1}T_{N(t)+1} \to 1$ w.p.1 as $t \to \infty$. Hence $\tilde{T}_1(z(\delta)/t) \sim t$ w.p.1 as $t \to \infty$. Thus the stopping rule is asymptotically independent of the scaling constant Γ. As a consequence, formation of asymptotically valid confidence intervals is impossible. In fact, even the asymptotic scaling of the rule is incorrect. It is well known that for estimation problems of the type described in Example 4.2.1, the amount of simulation time required to obtain an absolute precision of order ϵ is of order ϵ^{-2}, whereas the stopping rule $\tilde{T}_1(\epsilon)$ based on $\tilde{\Gamma}(t)$ in (2.75) yields a termination time of order ϵ^{-1}.